Grid y Computación de Altas Prestaciones

GRyCAP

Comparison of Container-based Virtualization Tools for HPC Platforms.

Diana M. Naranjo, Germán Moltó, Jorge Gomes, Mario David, Ignacio Blanquer

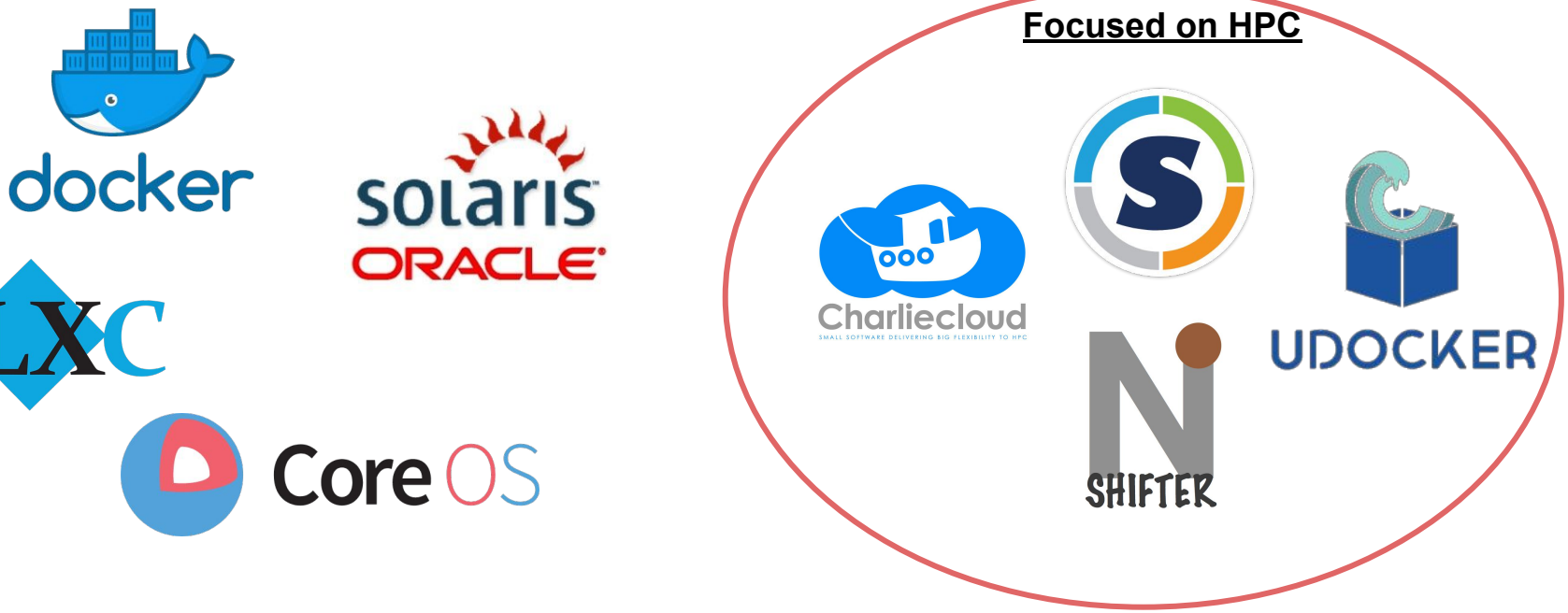
IBERGRID 2019
September 23th-26th Santiago de Compostela, España

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

i3M

CONTAINERS

- A standard unit of software that packages up code and all its dependencies.
- The application runs quickly and reliably from one computing environment to another.
- Interest for HPC: Users should only have access to their own data.



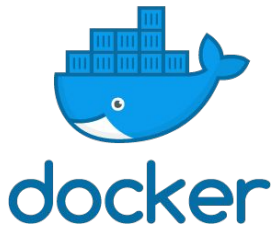
DOCKER

- Most popular container technology.
- Designed for microservices virtualization.
- Easy to deploy in a cloud, well documented, widely used by the developer community.

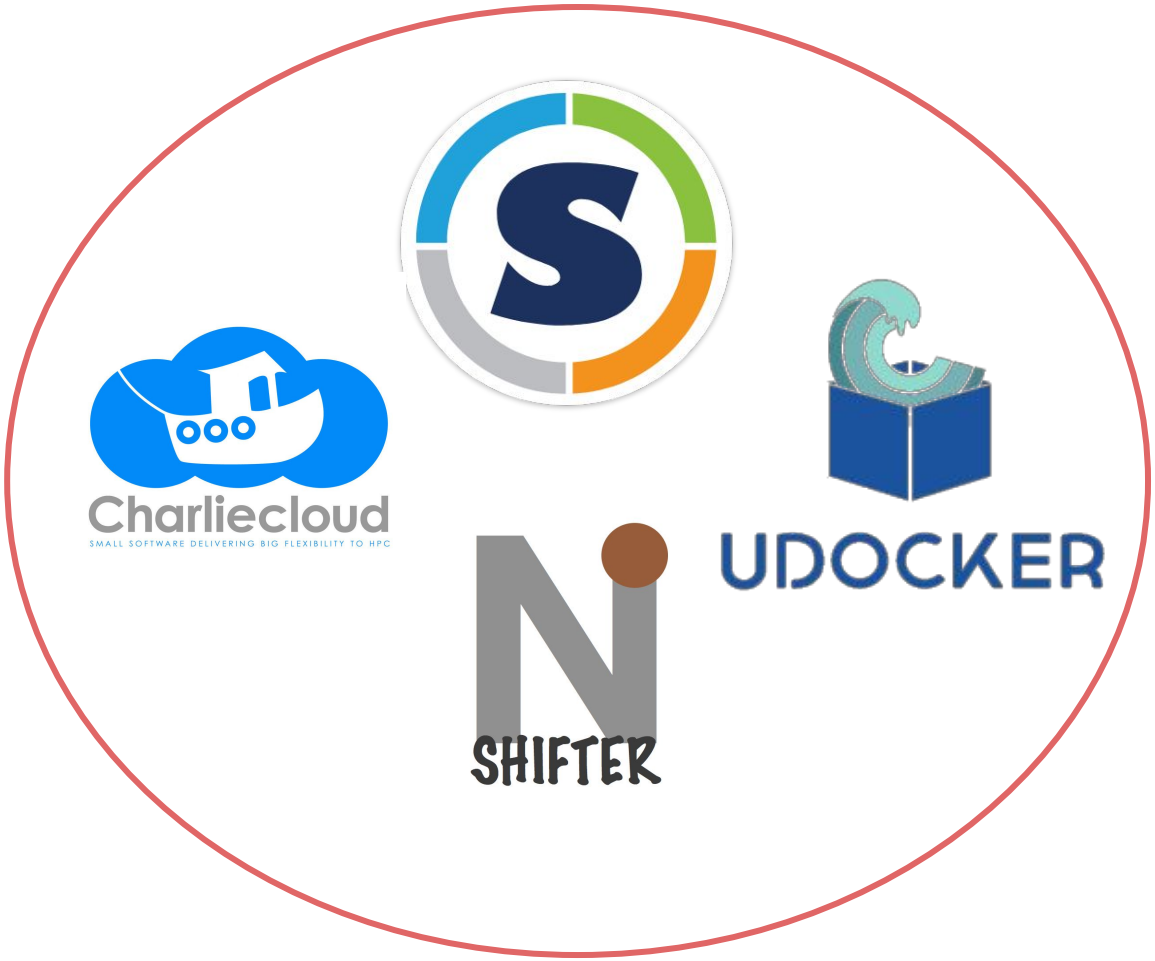
What about Docker in HPC?

- Users with access to the Docker socket can execute containers that may access system resources.
- Docker cannot safely use the host network stack and requires a separate network namespace.
- Docker instances are started by a separate daemon making integration with batch systems difficult including process control, accounting and resource usage.

HPC centers do not use it!!!

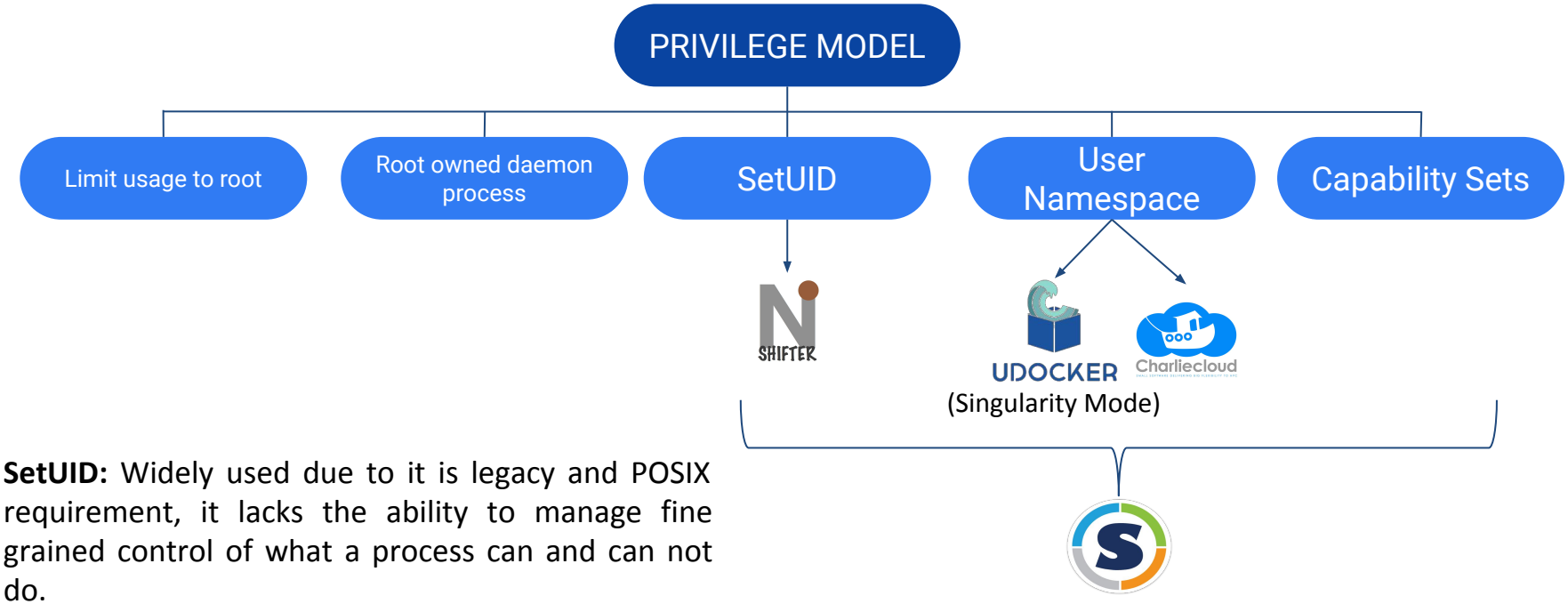


Container solutions oriented to HPC



Comparisons of different HPC orientated container solutions.

- Privilege model.
- Deployment.
- Interaction with Docker.
- Portability.
- Support MPI, InfiniBand and GPU.
- Security.



SetUID: Widely used due to it is legacy and POSIX requirement, it lacks the ability to manage fine grained control of what a process can and can not do.

User Namespace: The Linux kernel's user namespace may allow a user to virtually become another user and run a limited set privileged system functions.

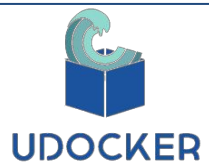
Capability Sets: Linux handles permissions, access, and roles via capability sets.

UDOCKER

-PRoot
-Fakechroot
-runC

chroot like environment over the extracted container.

DEPLOYMENTS



Just download and execute udocker and the installation will be performed automatically. (no root access required)

root required to install singularity, setuid need to work correctly.
Need Go installed in the host.

Download the GitHub repo and build.
Most commands require docker installed, so it requires root access.

Shifter need to be installed in all HPC nodes. Needs an Image Gateway (depends on MongoDB server, squashfs-tools, virtualenv and Python 2.7). Require root access to build and compile.

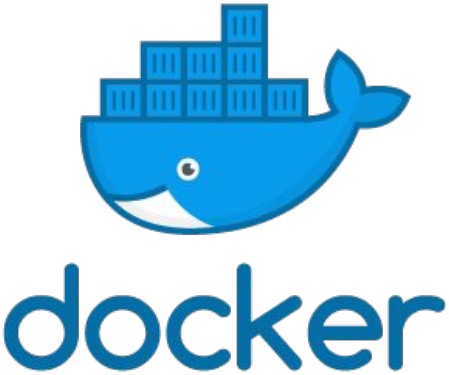
INTERACTION WITH DOCKER.

Does not use Docker at all. It has the ability to import Docker images.

Works completely independent of Docker. It has the ability to import Docker images, convert them to singularity images, or run Docker containers directly.

Needs Docker installed to run most of the commands. It can import Docker images.

Primary workflow is to pull and convert Docker images into Shifter images. Image Gateway connects to Docker Hub using its build-in functions; Docker does not need to be installed.



PORTABILITY

.sif format

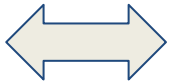


IMAGE
GATEWAY

SUPPORT FOR MPI AND INFINIBAND

	MPI	InfiniBand
UDOCKER	YES	YES
SINGULARITY	YES	YES
Charliecloud	YES	YES
SHIFTER	YES	YES

SUPPORT FOR GPU

UDOCKER

udocker setup --nvidia <container_name>
(add NVIDIA libraries and binaries)

--nv option to enable NVIDIA support.
(available with singularity exec, shell and run)

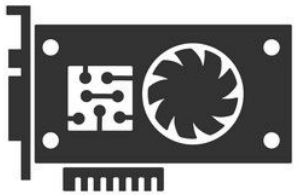
SINGULARITY

Charliecloud

./ch-fromhost --nvidia IMGDIR
(recommended by NVIDIA, via "nvidia-container-cli list")

File udiRoot.conf.
siteResources=/opt/shifter/site-resources (enable features like GPU support.)

SHIFTER



CHARACTERISTICS OF TEST ENVIRONMENT

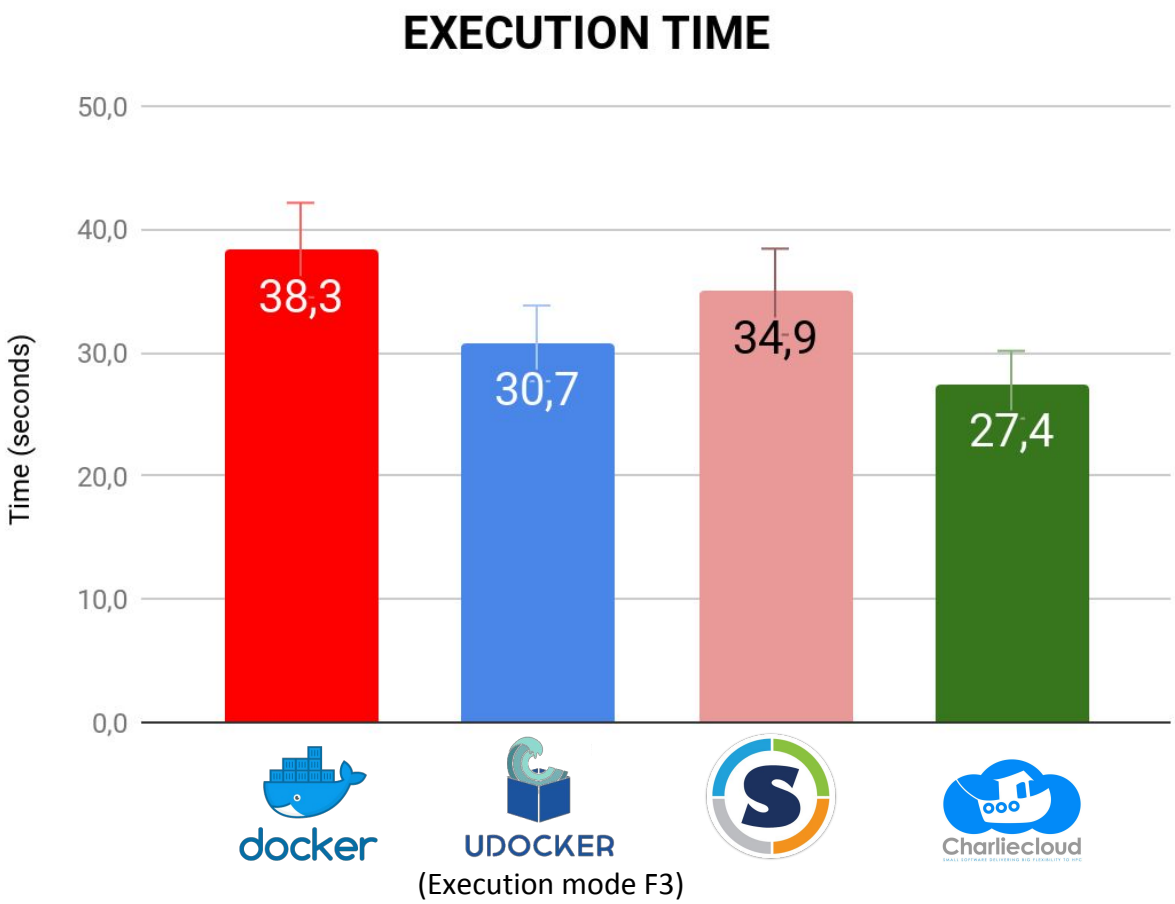
- Two nodes.
- The first one features two Skylake Gold 6130 at 2.1 GHz, 16 cores each, 768 GB RAM DDR4@2666, 10 GbE and includes an FPGA Arria 10 GX115 8GB, a RADEON Instinct MI25, 16GB, a Tesla P40 24GB and a Tesla V100 32GB.
- The second node features the same processor and memory together with 4 Tesla V100 32GB.
- Flavor Details: t2.large.nvidia-V100 (2 VCPUs, 8GB RAM, 40GB Size, 1 Tesla V100 32GB).

BENCHMARKING TO TEST THE GPU

Container:

- Latest GPU version of Tensorflow (from Docker Hub).
- Train a model to recognize handwritten digits (the MNIST data set).

https://github.com/tensorflow/models.git



(Execution mode F3)

SECURITY

- Due to the lack of isolation features udocker must not be run by privileged users.
- Files are extracted to the user home directory, normal file system protections apply.

- Root access to install singularity.
- Users run singularity image/app without special privileges.

- All the implications of having Docker installed.

- Root to install shifter binary.
- Need special integration into scheduler.

CONCLUSIONS

- Although Docker is the most popular container technology, it is difficult to apply to HPC clusters due to integration and security limitations.

- Of the container virtualization technologies for HPC centers, the most used today is Singularity.

- User experience:

  Singularity, Shifter and CharlieCloud are mostly oriented at being deployed by a system administrator.

  udocker is a user oriented tool and does not require system installation or privileges.

  CharlieCloud is very easy to use, but depends largely on Docker which requires privileged access.

  Shifter is very difficult to install in Ubuntu due to its many dependencies. Better prepared for CentOS.

CONTACT

Diana María Naranjo Delgado * email: dnaranjo@i3m.upv.es
Germán Moltó * email: gmolto@dsic.upv.es
Jorge Gomes ** email: jorge@lip.pt
Mario David ** email: david@lip.pt
Ignacio Blanquer * email: iblanque@dsic.upv.es

* Instituto de Instrumentación para Imagen Molecular
Universidad Politécnica de Valencia
Camino de Vera S/N, 46022, Valencia, Spain

** Laboratório de Instrumentação e Física Experimental de Partículas
Universidade de Lisboa
Av. Prof. Gama Pinto 2, 1649-003, Lisboa, Portugal